

Remarks

As stated above, the applicants appreciate the examiner's thorough examination of the subject application and request reexamination and reconsideration of the subject application in view of the preceding amendments and the following remarks. Claims 1-24 remain in the application. All of the claims have been amended in order to clarify the original claim language.

Claims 1, 11, 21 and 22 have been rejected under 35 U.S.C. §112, first paragraph, as failing to comply with the written requirement. Claims 1-7 and 11-17 have been rejected under 35 U.S.C. §103(a) as being unpatentable over Matthews et al. (the '807 Patent) in view of Bartholomew et al. (the '119 Patent). Claims 8, 10, 18 and 20 have been rejected under 35 U.S.C. §103(a) as being unpatentable over Matthews et al. in view of Bartholomew et al. and further in view of Miner et al. (the '789 Patent). Claims 9 and 19 have been rejected under 35 U.S.C. §103(a) as being unpatentable over Matthews et al. taken in view of Bartholomew et al. and further in view of Szlam et al. (the '731 Patent). Claim 21 has been rejected under 35 U.S.C. §103(a) as being unpatentable over Matthews et al. taken in view of Bartholomew et al. and further in view of Miner et al. and Szlam et al. Claims 23 and 24 have been rejected under 35 U.S.C. §103(a) as being unpatentable over Matthews et al. in view of Bartholomew et al. and further in view of Szlam et al. and Brown et al. (the '180 Patent). These rejections are respectfully traversed and reconsideration is requested in view of the foregoing amendments and following remarks.

All of the claims have been amended to clarify the claimed subject matter, or to make grammatical corrections.

With respect to the rejection of claims 1, 11, 21 and 22 under 35 U.S.C. §112, first paragraph, as failing to comply with the written requirement, the claims have been amended to delete the phrase “for whom a voice pattern template is not defined”, which was added by the previous amendment, and which was the basis of this rejection. By deleting this limitation, the claims are placed in the original form, although the claims have been further amended to refer to the “speech recognition analysis”, “speech recognition application”, “speech recognition device”, all as “speaker-independent”, to determine the semantic meaning of a spoken response provided by an answering person (as opposed to the verification of the identity of such person). These terms have been inserted to make clear the distinction between “speech recognition” (SR) and “voice verification” (VV), in the context of the present application. The former term, “speaker-

independent”, means that the analysis, application or device responds regardless of the speaker. The latter term, “voice verification” means that the analysis, application or device has been previously “trained” by the speaker (by repeating the desired utterances to the application or device several times so it may optimize itself to the speaker’s unique vocal characteristics, i.e., to the speaker’s “voice template”) to verify the identity of the speaker (by comparing a given utterance to the voice template), and is not adapted to recognize the meaning or identity of anyone else speaking, other than the person whose voice template has been stored in the system. Stated another way, one should not confuse speech recognition, where signal analysis by a computer is used to identify the semantic meaning of what any human says, in a speaker-independent fashion, even when the system has never heard the person’s voice before, and voice verification, where signal analysis is used to determine if someone’s voice matches pre-existing samples/templates in order to identify/verify who that person is. While the term “speaker-independent” has not been expressly stated in the specification of the application, it can clearly be inferred from a reading of the present application. Specifically, the various trigger events described in the application occur in response to the person answering the telephone regardless of who answers. It is also very clear from the specification that the system analyzes the semantic meaning of responses of the person on the line regardless of who is speaking.

Referring to the rejections based upon prior art, it is submitted that the cited references do not anticipate nor make obvious applicants’ invention. The present application describes a method of and system for controlling an outbound call, using speaker-independent speech recognition (SR) to identify the semantic meaning of what the answering party has said. The application describes other methods of detecting answering machines, busy lines, etc. Thus, for example: one can place a call and identify "Hello" (or "Doe Residence", or a whole variety of other greetings one may include within the system) when a person picks up the telephone. One can then play "Hello, this is Joe calling for John Doe - is he available?" The answering party (independent of who it is) can then respond with one of literally thousands of potential responses (such as Yes, that's me, Yes I am, Sure go ahead, No he's not, Can you hold on a minute, Can I take a message....etc.). Of significance, one does not need or have any experience with the answering party’s voice or "templates", and specifically does not need training with the answering party's voice, and the interaction (via logical branching) of the pre-recorded prompts and flexible speech recognition creates a simulated conversation between the person and the

computer. The variety and combinations and complexity of the responses one can handle are huge and by no means "obvious." It is submitted that none of the cited references either anticipate or make obvious applicants' invention.

Matthews describes a Voice Message System (VMS) for the deposit, storage, and delivery of audio (i.e., voicemail) messages. A user can deposit recorded messages in the system, and instruct it to deliver these messages to other addresses/extensions on the system. The user can also check the system for recorded messages left for him/her by other users. To gain access to the system the user must key in (touch tone) an access code and/or say a selected password (with the password being compared to an "associated distinctive voice feature template", i.e., voice verification to confirm the identity of the user). When the VMS does outbound calls to deliver a recorded message, it can enable a Name Announce Feature, e.g., "This is VMS." There is a voice message for [name]." Then the user must key in his ID (touch tone) and VMS simply plays the recorded message. There is no mention of speaker-independent speech recognition, nor interaction/branching, and certainly not based on understanding the semantic meaning of what the user/recipient says.

It is submitted that none of the other cited references overcomes the deficiencies of the Matthews patent. Bartholomew describes a method of and system for providing personalized calling services. The system and method enable a user to load and change service features into a phone switch. For example, a user could direct that his or her calls made from a hotel be billed to his or her business number, roommates on the same number could have different ring tones and billing, parents could require callers to verify their identity when calling into their home when children are home alone (or require each child to verify the identity of the caller and then control who he or she can call accordingly), or to identify and/or screen harassing calls. Users access the system via voice verification. Non-users calling a personalized line are identified using voice verification as well. The reference makes many references to voice templates and "speech based identification" - i.e., voice verification. Note how the spec defines that callers must be instructed specifically on what to say (e.g., "This is Jane") - this is so the voice verification will work. Bartholomew describes one embodiment in column 45 lines 40-65, in which a resident of the premises other than the harassed party calls the harassed party "in such a manner" that his call for the harassed party is fully audible to the telephone microphone and is

identified by the bridged IP SIV, and the CPR of the harassed party is substituted by the destination central office switch.. It is submitted that this description must be taken in context:

In the foregoing example it was assumed that the harassed party answered the telephone. The profile of that person was then installed based on that person identifying his or herself in answering the telephone. In the case where the telephone is answered by a resident of the premises other than the harassed party, that party answers with the same pre-specified greeting, such as, "This is John." In this situation the profile in the CPR corresponding to the subscriber line may specify that the profile of that individual, namely John, be substituted for the line CPR. Unless the caller specifically requests to speak to the harassed party by name, the call will proceed in accord with the CPR of the answering party, in this instance, John.

If the caller, as yet unidentified, asks to speak to the harassed party, Jane in this example, alternative procedures may be utilized according to this embodiment of the invention. According to a first procedure, the answering party, John, calls the harassed party, Jane, in such a manner that his call for Jane is fully audible to the telephone microphone. This utterance is identified by the bridged IP SIV (intelligent peripheral speaker identification verification), and the CPR of Jane is substituted by the destination central office switch. Jane answers in the pre-agreed format, such as, "This is Jane." The CPR for Jane has been entered and the same procedure is followed as has been previously described in the instance in which Jane answered the telephone.

As an alternative to this procedure, an answering party other than Jane may say "Please hold," place the call on hold, and call Jane to the telephone. Jane may then remove the call from hold, and answer in the pre-agreed manner, such as, "This is Jane." The CPRs of all residents of the subscriber premises may contain call handling instruction data directing interpretation of hold signals on this line as directing connection to the SIV facilities of the IP and directing the IP to stand by to implement CPR selection corresponding to the next matching name of a subscriber. As a result the switch substitutes the CPR for Jane on identifying her name through SIV. The CPR for Jane is now entered and the same procedure is followed as has been previously described in the instance in which Jane answered the telephone. As a still further alternative to the foregoing, all subscriber OE profiles at the subscriber premises may contain processing instructions to cause the IP to be connected upon execution of a *HOLD sequence. In this case all subscribers are instructed to use a *HOLD sequence when calling the threatened party to the telephone.

Thus, the only speech recognition used, if at all, is to substitute one voice verification template for the person answering the telephone for another based solely on the response of the resident answering the phone in some preconceived manner. It appears that speaker-independent speech

recognition is not used in Bartholomew for the various features set forth in applicants' claims. Assuming, *arguendo*, that Bartholomew can be construed to suggest that the system and method described by Bartholomew does suggest speech recognition because some interpretation of semantic meaning is performed during the process, it is clear that the speech recognition is "speaker-dependent", and not "speaker-independent". Speaker-dependent speech recognition systems, like voice verification systems, must be trained by the user (whose voice is to be recognized). The user trains the speech recognition system or application to recognize the semantic meaning of certain phrases (as opposed to the speaker's identity) with enhanced accuracy, and thus must be trained by the intended user, in a similar manner to that of voice verification. A speaker-dependent speech recognition system or application will, with a high degree of probability, not understand someone who has not trained the system. A speaker-independent system or application on the other hand will recognize the semantic meaning of what is said by persons whom the system and application have never heard.

Applicants are submitting by way of an information disclosure statement four background references: Ed. Cole, Ronald A., et al., "Survey of the State of the Art in Human Language Technology", Sponsored by the National Science Foundation, et al., November 21, 1995, table of contents, and pages 11-70; Cox, Richard V., et al. "Speech and Language Processing for Next Millennium Communications Services", Proceedings of the IEEE, Vol. 88, No. 8 (August 2000), pp 1314-1337; Zue, Victor W., et al., "Conversational Interfaces: Advances and Challenges", Proceedings of the IEEE, Vol. 88, No. 8, August 2000, pp. 1166-1180; Foster, Peter, et al., "Speech Recognition, The Complete Practical Reference Guide, published by Telecom Library, Inc., August, 1993, table of contents, pp.1, 38 and 39. These references are being provided to aid the Examiner in his understanding of the differences between speaker-dependent and speaker-independent speech recognition, as well as speaker-dependent voice verification. Applicants are also willing to submit the declaration of a person skilled in the art to further support this understanding.

Miner describes a computer-based assistant ("Wildfire") for screening and managing inbound calls. When someone calls a subscriber, the system asks the caller to state their name. It then records the name and uses a speaker-dependent dictionary/contact file (i.e., names must be pre-registered) to attempt to recognize the caller. If not recognized, the caller must key in their phone number. The subscriber can then choose to take the call, put it on hold, into

voicemail, automatic call back later, etc. The specification of Miner is quite clear that voice templates or registrations are needed for the names, and thus the system must be speaker trained, and therefore is speaker dependent. Again, assuming, *arguendo*, Miner can be construed as teach speech recognition because the system does interpret the semantic meaning of certain speech, it is clear that such speech recognition would be speaker-dependent, as opposed to speaker-independent. It appears speaker-independent speech recognition is not used in Miner for the various features set forth in applicants' claims.

Szlam describes a method of and system for telemarketers to minimize the annoyance of their hapless targets. When a telemarketer's automated dialer gets someone on the phone (there appears to be no mention of speech recognition, so one must assume that the system detects that someone is on the line by telecom signaling of a handset pick up or simple detection of volume level on the line), and an agent is not available to talk to them, it plays an apology message. There is also mention of also apologizing for wrong numbers, but the patent is silent as to how this determined (it appears that this may be a ruse to minimize the annoyance of answering parties who would otherwise hear silence on the line when agents are unavailable). The patent reference also describes using the level and duration of a signal to detect an answering machine. Once again, it appears that the reference makes no mention of a speaker-independent speech recognition system with the various features set forth in applicants' claims.

Brown describes a message delivery system for selecting the language to be used in the system announcement before each message delivery (e.g., for international use). There is no mention of speaker-independent speech recognition. The only language that the Examiner references is having a user employ touch tone entries to interrupt/skip an instructional prompt. The Examiner relates this to how the present applicants use a beep to interrupt a turn-taking approach used for leaving an appropriately timed message on an answering machine-this is an unrelated function performed by the application or system, not the user.

Referring to the Examiner's specific arguments by paragraph number of the outstanding Official action:

Paragraph 2. For one feature, Matthews does indeed teach a call to a telephone station/number that lists a target person for whom a voice pattern template is not defined (col 83, lines 30-34). But for the feature referenced ("Name Announce" - for delivery of voicemail messages with the recipient's name included in the greeting) Matthews just plays the called

party's name and asks them to key in their ID via touch tone in order to hear the message ("Hello, this is VMS. There is a message for [name]. Please dial your I.D."). No speaker-independent SR, much less voice verification is used with this feature.

Paragraphs 3 and 4. The rejection under 35 U.S.C. § 112 is addressed above. In addition, the examiner objects that the limitation added in the last response ("a target person for whom a voice pattern template is not defined") was not in our original specification. This was added to distinguish speaker-independent speech recognition from voice verification to clarify these two fundamentally different technologies, based on the objections in the first office action. The claims have now been modified to make the same distinction. The basis for the amendments have clear support in the present application. By definition, since the present system is used for outbound calling, it will frequently contact households that have not been called before (and even if the household was called before, a different person may answer), it is clear that one can not control who is reached at each household. The system can not be pre-supplied with pre-defined voice templates. They are not needed. The system recognizes what the person says, not verifying who they are. So, once again, the Examiner should take notice of the distinct difference between speaker-independent speech recognition and voice verification.

6. While Matthews does teach a system for making an outbound call placed to a target person for whom a voice pattern template is not defined, he makes no mention of capturing a spoken response at all, or more specifically of using speech recognition to identify the meaning of what the answering party says (in order to select the next prompt and provide a turn-taking interaction). In the reference used, he simply plays the name and specifically states that the ID must be keyed in ("This is VMS. There is a message for [name]. Please dial your ID." No mention is made of speech recognition). Bartholomew does capture a spoken response, but for voice verification to confirm the identity of the answering party. Whether it is obvious to "modify Matthews to allow speech recognition analysis to determine the target person as taught by Bartholomew" is irrelevant, since this is not what is required by the pending claims. The pending claims do not determine/verify/identify who the target person is. The pending claims recognize/identify the meaning of what a person says. The Examiner also references Bartholomew for "monitoring the conversation" after the target's identity is verified using voice verification. The presently claimed system and method don't require monitoring of the conversation as a third party/second link, but rather interact directly with the target by

determining the semantic meaning of what they said and selecting the next prompt accordingly (i.e., passive monitoring vs. interacting via speech). Regarding claims 2 and 12, Bartholomew uses voice verification (i.e., confirmation of a person's identity via their unique vocal characteristics) to determine when an answering party is not the target, not speaker-independent speech recognition to determine the semantic meaning of what they said. Speech recognition requires dictionaries, grammars, large sets of allowed responses to do this, not voice verification.

7. Present claims 8 and 18 involve recognizing during an out bound call when a called party asks "Who is calling?" (and its myriad variations) and then playing a pre-recorded response indicating who the calling party is. So for example when calling someone during an outbound call, the system recognizes "Who is calling?" and responds with "It's John calling." As the examiner notes, neither Matthews nor Bartholomew teach this. Miner teaches asking an INBOUND caller to state his or her name (specifically) so that the name can be recognized if it is part of a speaker-dependent dictionary of allowed contact names. Miner includes instructions for repeating names so that they can be added to the dictionary. If not recognized the name is recorded and the caller must key in their phone number. If the number is recognized by the subscriber's system a recording is played to ask the caller to provide a voice template by stating their full name two more times. Miner uses this method to screen/route the inbound call (to the called party, to voicemail, to be put on hold), as a third party/secondary link between the caller and the called party. Not recognizing all the speaker independent responses, Miner requires registering/a template for the name to be recognized.

Claims 10 and 20 involve knowing when one has not recognized the meaning of a spoken response (e.g., due to low confidence thresholds), and repeating a pre-recorded greeting and asking for the target person again. Matthews and Bartholomew fail to teach this. Miner teaches having a subscriber train the system on his specific voice, giving specific commands - so that the subscriber can control inbound calls (e.g., put to voicemail, put on hold, connect when done with current call). This is speaker-dependent speech recognition used for an inbound call. Using the presently claimed system one cannot control and does not know who will answer the outbound call or what they will say. The voice characteristics are not known prior to the call, and the responses are much more varied. Modifying Matthews non-speech voicemail delivery system, with Bartholomew's voice verification user identification, with Miner's speaker-dependent

commands for control of inbound calls....is not an obvious link to outbound call control to a target without a defined voice template, i.e. speaker-independent speech recognition.

8. As the examiner notes, Matthews and Bartholomew do not teach detecting via speech recognition that a wrong number was reached and then playing an apology message. But neither does Szlam - see col 2 lines 51-58 that the examiner references. What Szlam discloses is that if an out-bound call is answered, and there is no live agent available, then just play a message that apologizes for the inconvenience or indicates that a mistake, SUCH AS A WRONG NUMBER, was made - regardless. If the call is answered and no agent is available they just apologize for dialing a wrong number whether or not it is in fact wrong - to hide the purpose of the call/lack of agent availability. There is no disclosure for how Szlam would detect a wrong number.

9. As discussed above, Matthews uses voice pattern templates (voice verification) to confirm the identity of users who want access to the VMS to deposit or retrieve voicemail messages. The one case where a call is placed without a voice template is to deliver a message to an extension selected by a user. In this case one doesn't receive a spoken response to verify the recipient at all - the system just says "There is a message for [name]. Please dial your ID" (touch tone entry). Bartholomew teaches capturing a spoken response, but using a template for voice verification. Again, the presently claimed system does not determine the identity of the target person as taught by Bartholomew - it identifies what a person says without confirming who they are. "Monitoring" a call as a secondary link in order to modify service profiles is not the same as interacting with the called party directly via SR and prompts that branch depending upon the meaning of what the person said. As is understood in the art, voice verification or authentication is not speech recognition. Regarding the claim limitation 21(E)(c), leaving a message with a non-target person is not the same as having a calling party record a message for delivery per Bartholomew. Regarding limitation 21(E)(d), the same argument applies for voice verification as compared to speech recognition. Regarding limitation 21(E) (e), Miner recognizes the name of an inbound caller to route the call to the subscriber (if the name has a template that has been registered/trained in the system). The present system and method recognizes "Who is calling?" on an outbound call and states who the calling party is. For limitation 21(E)(f), Szlam does not teach that a wrong number was made, he just plays this type of apology message as a ruse to cover the true purpose of the call - and neither Matthews or

Bartholomew disclose recognizing when a person says that the system has a wrong number (all VV, no SR).

Regarding 21(E)(g), repeating the request for a target because an utterance was not recognized on an outbound call is not related to repeating Miner's prompts when a subscriber resumes accepting inbound calls.

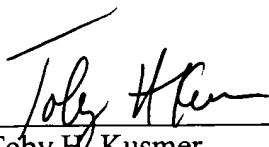
10. Answering machine detection: The predetermined time period claimed by applicants is not a hold state per Bartholomew, but a time limit to initiate the "turn taking" rules that determine whether a human or answering machine has answered the call. The present system and method are looking for speech over a certain duration without a pause, even while a prompt is playing (a human usually stops to take turns in conversation). This is not to be confused with a hold time, e.g., when a non-target goes to get the target on the phone. Matthew's name announcement feature is discussed above – the user ID must be keyed in to hear the message.

11. Beep detection: Brown teaches tone prompting to start the recording of a message ("start recording your reply at the beep"). The Examiner says Brown also "inherently" teaches interrupting the announcement upon detecting a tone. This is not understood since applicants believe no such inherency exists.

No new matter has been added by these amendments. The applicants respectfully assert that the subject application is now in condition for allowance. Please apply any charges or credits to deposit account 50-1133.

Respectfully submitted,

Date: 7.13.04



Toby H. Kusmer
Reg. No. 26,418
McDermott, Will & Emery
28 State Street
Boston, MA 02109
DD: (617) 535-4065
Fax: (617) 535-3800
E-mail: tkusmer@mwe.com